Project Title	Prototype of HPC/Data Infrastructure for On-demand Services
Project Acronym	PHIDIAS
Grant Agreement No.	INEA/CEF/ICT/A2018/1810854
Start Date of Project	01.09.2019
Duration of Project	36 Months
Project Website	www.phidias-hpc.eu

D1.7 - First report on scientific committee

Work Package	**WP 1, Management**
Lead Author (Org)	**Florian PIFFET & Corentin LEFEVRE (Neovia Innovation)**
Contributing Author(s) (Org)	**François BODIN (Rennes 1 University), Thierry BIDOT (Neovia Innovation)**
Due Date	**01.08.2020**
Date	**DD.MM.YYY**
Version	**V7.0**

Dissemination Level

☒ PU: Public

□ PP: Restricted to other programme participants (including the Commission)

□ RE: Restricted to a group specified by the consortium (including the Commission)

□ CO: Confidential, only for members of the consortium (including the Commission)

Co-financed by the Connecting Europe Facility of the European Union

The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

Versioning and contribution history

Version	Date	Author	Notes
1.0	07.07.2020	BIDOT Thierry (Neovia Innovation), PIFFET Florian (Neovia Innovation)	First draft
2.0	09.07.2020	LEFEVRE Corentin (Neovia Innovation)	Second draft
3.0	15.07.2020	BODIN François (Université Rennes 1 – SUC Coordinator)	Third draft
4.0	17.07.2020	LEFEVRE Corentin (Neovia Innovation)	Final draft
5.0	29.07.2020	LEFEVRE Corentin (Neovia Innovation)	Additional information on SUC presidency
6.0	04.08.2020	SUDRE Joël (CNRS), Jean-Christophe PENALVA (CINES)	Reviewed version
7.0	12.08.2020	GERMETZ Emilie (Neovia Innovation)	Final editing

Disclaimer

This document contains information which is proprietary to the PHIDIAS Consortium. Neither this document nor the information contained herein shall be used, duplicated or communicated by any means to a third party, in whole or parts, except with the prior consent of the PHIDIAS Consortium.

Co-financed by the Connecting Europe Facility of the European Union

The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

Table of contents

The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

Terminology

Item	Description
AI	Artificial Intelligence
CA	Consortium Agreement
CIM	Common Interface Model
CMR	Common Metadata Repository
DAO	Data Access Object
DIAS	Data and Information Access Services
DIVAnd	Data Interpolating Variational Analysis N Dimensions
EC	European Commission
EMEA	Europe, Middle East and Africa
EO	Earth Observation
EOSC	European Open Science Cloud
ESGF	Earth System Grid Federation
EU	European Union
HPC	High-Performance Computing
HPDA	High-Performance Data Analytics
HS	Human Sciences
HTC	High-Throughput Computing
ICT	Information and Communications technologies
KPI	Key Performance Indicator
ML	Machine Learning
OGC	Open Geospatial Consortium
OSU	Observatory for Universe Sciences
PCA	Principal Component Analysis
PHIDIAS	Prototype of HPC/Data Infrastructure for On-demand Services
PM	Particulate Matter
PMO	Project Management Office
S5P	Sentinel 5 Precursor
STAC	Spatio Temporal Asset Catalog
SUC	Scientific and Users Committee
WP	Work Package

Co-financed by the Connecting Europe Facility of the European Union

The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

Executive Summary

The European project PHIDIAS aims to build a prototype for Data/High Performance Computing services based on Earth sciences cases. The consortium will develop and provide new services to discover, manage and process spatial and environmental data produced by research communities tackling scientific challenges such as atmospheric, marine and earth observation issues.

Along with the Management Board, the Technical Board and the Security Committee, the Scientific and Users Committee (SUC) is one of the main bodies in PHIDIAS, specifically dedicated to provide guidance on technical challenges and give advices on potential solutions and choices, as well as it may help connecting to global organizations and communities.

This Commitee gathers several experts external to the PHIDIAS project that will meet periodically to fulfil their missions.

This document reports the activities conducted during the first 11 months of the project (September 2019 to July 2020) first to establish the project's SUC and to cover the first works done by the external experts.

Co-financed by the Connecting Europe Facility of the European Union
The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

1 Introduction

PHIDIAS is a project addressing complex challenges with different stakeholders in a fast-changing environment.
For these reasons, it has been decided since the beginning to create a Scientific and Users Committee to provide guidance on technical challenges and give advices on potential solutions and choices. In addition, it may help connecting to global organizations and communities.

The establishment and running of such a committee is taken in charge under T1.3 of the project.
During the first months of PHIDIAS, the focus has been to create such a Committee and gather experts from various fields and scientific communities.
In PHIDIAS, the SUC is organized around three roles:

- A president who decides meeting occurrences, elaborate meeting agendas and chair meetings and discussions
- A coordinator who implements meeting logistics and write minutes, interacts with project participants, collects project data and information, disseminates recommendations, and supports the president in his/her role
- External experts who participate to meetings and discussions related to the project, challenge the project's assumptions, plans and choices and provide guidance and advices

PHIDIAS' Project Management Office (PMO) supports the implementation and the work of the Committee.

Co-financed by the Connecting Europe Facility of the European Union
The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

2 Scientific and Users Committee’s organization

2.1 Implementation

The PHIDIAS Management Board in coordination with the Technical Board designated the participants (external experts) of the SUC based on its proposals and the ones from François Bodin (Rennes 1 University), Thierry Bidot (Neovia Innovation) and Boris Dintrans (PHIDIAS Coordinator).

2.2 Structure and rules

The role and the organization of this the Scientific and Users Committee are the following.

The Committee have a general guidance role in the PHIDIAS project. This includes providing guidance and opinion on project technical challenges, solutions and choices, including the project technical architecture. Furthermore, due to its experts’ composition, the SUC may enable connections to world-wide groups and organizations.

The Committee President, nominated among its peers, oversees the meeting occurrences and elaborates meeting agendas. He chairs meetings and discussions. François Bodin (SUC coordinator – see below) has been elected president of the project’s SUC on 28 July 2020.

The Committee Coordinator (François Bodin), as the main support of the President, is in charge of the good management of the SUC: he implements meeting logistics and writes minutes, interacts with project participants and collects project data and relevant information to disseminate recommendations.

Co-financed by the Connecting Europe Facility of the European Union

The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

2.3 List of Scientific and Users Committee's external experts

The following parameters have been applied by the Management Board when choosing the external experts for the SUC:

- Recognition at international level
- As a whole, the experts shall cover all scientific and technical challenges of the project
- Participation to other major projects in the field
- Gender balanced
- As a whole, the SUC shall not comprise more than 10 experts to ensure that the exchanges are fluid and dynamic

Experts have been proposed by the partners and pre-selected following the above-mentioned criteria. Below is the current list of the experts who accepted to enter and participate to the PHIDIAS' SUC.

Jean-Thomas Acquaviva (FR) holds a PhD from the University of Versailles performed at CEA/DAM and is a Team lead Research Group at DDN Storage. He formerly worked for Intel, the University of Versailles and the French Atomic Commission (CEA), where he participated to the creation of their joint laboratory the Exascale Research Centre. He is also the Coordinator of the H2020 Evolve project to tackle Big Data processing.

Publications: https://dblp.org/pers/a/Acquaviva:Jean=Thomas.html

François Bodin (FR) held various research positions at University of Rennes I (where he holds a Professor position and the chair "Mobility in a sustainable city" awarded by the Rennes 1 Foundation) and at the INRIA research lab. His contribution includes new approaches for exploiting high performance processors in scientific computing and in embedded applications. He is the CEF AQMO Project Coordinator and hold the position of PHIDIAS' SUC Coordinator and President.

Publications: http://people.irisa.fr/Francois.Bodin/?page_id=2

Michele Fichaut (FR) is a PhD database engineer working at the Scientific Information Systems Service for the Sea (PDG-IRSI-SISMER) of the French Research Institute for the Exploitation of the Sea (Ifremer). Among her works in the Institute, she coordinated the FP7 project SeaDataNet II (2011-2015) and is currently the coordinator of the H2020 project SeaDataCloud, which is a key infrastructure adopting Cloud and HPC technology for better performance to drive several portals of the European Marine Observation and Data network (EMODnet), initiated by EU DG-MARE for Marine Knowledge, MSFD, and Blue Growth.

Publications: https://www.researchgate.net/profile/Michele_Fichaut

Maryvonne Gerin (FR) is an astrophysicist who works on the structure of the interstellar environment and its content (simple and complex molecules). She is the Research Director at the Radiation and Matter Studies Laboratory in Astrophysics (LERMA) of the French National Centre for Scientific Research (CNRS) and Assistant Scientific Director in charge of Research Infrastructures at CNRS/INSU. She notably participated in the Herschel space mission.

Publications: https://www.researchgate.net/profile/Maryvonne_Gerin

Research Director at the CNRS, Sylvie Joussaume (FR) is a climatologist specialised in the climate change. She is also the Director of the Paris Research Consortium Climate-Environment-Society which gathers the research capacity of 17 French laboratories working mainly in the fields of climatology, hydrology, ecology, health, and the humanities and social sciences. She received the 2017 Trophy of the Stars of Europe as coordinator of the H2020 project IS-ENES2 federating the European Community of Climate Modelers.

Publications: https://www.researchgate.net/scientific-contributions/2067006330_Sylvie_Joussaume

The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

Bruno Raffin (FR) is a Research Director at the INRIA, Head of the DataMove Research Team. His current research activities are focused on Data Analysis for High Performance Computing. He was responsible for INRIA of more than 15 national and European projects. Bruno Raffin has been involved in more than 30 program committees of international conferences. He is the head of the steering committee of the Eurographics Symposium on Parallel Graphics and Visualisation.

Publications: http://datamove.imag.fr/bruno.raffin/publications.php

Debora Testi (ITA) has an electronic engineer and a bioengineer degree from the University of Bologna. Most of her publications have been related with computer-aided software for chirurgical operations (hip replacement, prediction of femoral neck fractures, etc.). She is currently working in the HPC department (Middleware and Data Management group) of CINECA as project manager, notably on EU projects in ICT for Health and HPC infrastructures. Debora Testi is also a member of the Board of Directors of PRACE.

Publications: https://dblp.org/pers/t/Testi:Debora.html

Jean-Pierre Vilotte (FR) is a professor at the Institut de Physique du Globe de Paris and Director of the Parallel Computing and Data Analysis Centre of the Institut. His main research concerns earthquake seismology and wave tomography, parallel computing and data-intensive architectures, etc.

Publications: https://www.researchgate.net/profile/Jean-Pierre_Vilotte

2.4 *Process to select the SUC Coordinator*

A specific process has been conducted to select the SUC Coordinator that is fully described in annex 4.3 to this document.

Co-financed by the Connecting Europe Facility of the European Union

The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

3 Contributions of the SUC during the project’s first year

Contributions of the PHIDIAS’ SUC is organized through exchanges between the external experts and the project’s team, especially the Technical Board.

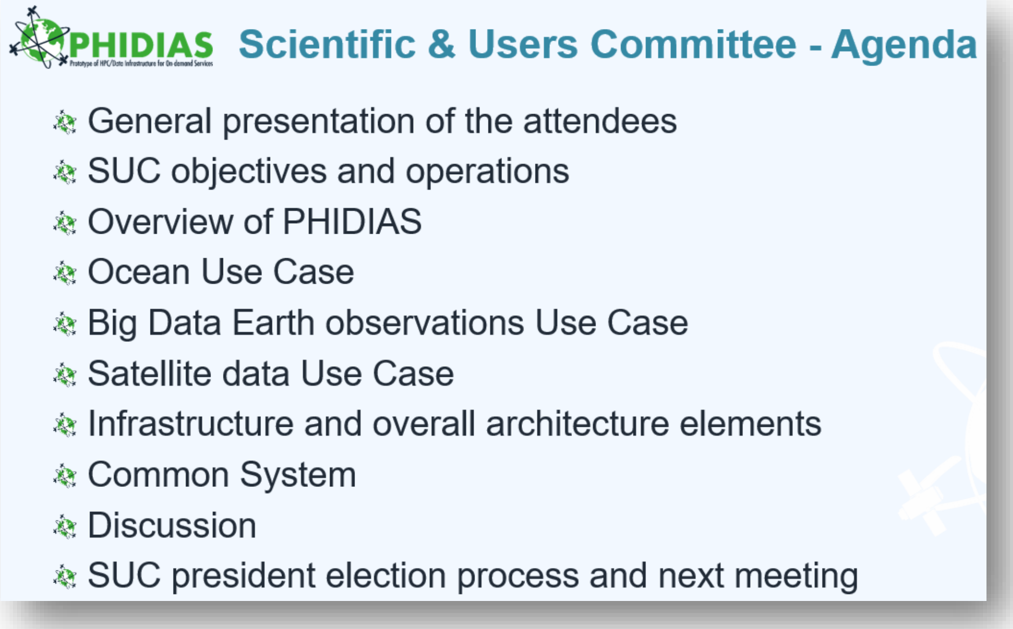
Initially a face-to-face meeting was planned in spring 2020 but due to the sanitary situation, this meeting has been changed to a remote meeting at the beginning of July, 2020, that gathered the external experts and the Technical Board members.

3.1 *Agenda of the SUC first meeting*

The agenda of the first meeting has been organized around the presentations of each participant, to connect the external experts with the project’s technical leaders.

The technical leaders had each a slot to present the current status of their work and the questions to be solved, with a moment of open dialogue for each project’s component.

The meeting ended with an overview of the transversal questions PHIDIAS has to tackle.



PHIDIAS Scientific & Users Committee - Agenda

- General presentation of the attendees
- SUC objectives and operations
- Overview of PHIDIAS
- Ocean Use Case
- Big Data Earth observations Use Case
- Satellite data Use Case
- Infrastructure and overall architecture elements
- Common System
- Discussion
- SUC president election process and next meeting

Figure 1 - PHIDIAS SUC first meeting agenda

3.2 *Minutes of the SUC first meeting*

3.2.1 *General presentation of the attendees*

Attendees	
Corentin LEFEVRE	Neovia Innovation, PMO/WP1
François BODIN	Rennes 1 University, PHIDIAS’ SUC Coordinator
Florian PIFFET	Neovia Innovation, PMO/WP1
Thierry BIDOT	Neovia Innovation, PMO/WP1

Attendees	
Boris DINTRANS	CINES, PHIDIAS Coordinator/WP1 leader
Jean-Christophe PENALVA	CINES, PHIDIAS WP2 leader
Cécile NYS	Ifremer, WP6
Pascal PRUNET	SPASCIA, PHIDIAS WP4 leader
Charles TROUPIN	University of Liege
Emilie GERMETZ	Neovia Innovation, PMO/WP1
Marion LEPAYTRE	CINES, PMO/WP1
Sylvie JOUSSAUME	CNRS, SUC member
Jean-Pierre VILOTTE	Institut de Physique du Globe, SUC member
Michèle FICHAUX	Ifremer, SUC member
Debora TESTI	CINECA, SUC member
Maryvonne GERIN	CNRS, SUC member
Jean-Thomas ACQUAVIVA	DDN Storage, SUC member
Bruno RAFFIN	INRIA, SUC member
Gilbert MAUDIRE	Ifremer, PHIDIAS WP6 leader
Frédéric HUYNH	CNRS, PHIDIAS WP3
Jean-Christophe DESCONNETS	IRD, PHIDIAS WP5 leader
Richard MORENO	CNES, PHIDIAS WP3 leader

3.2.2 SUC objectives and operations

The Committee have a general guidance role in the PHIDIAS project. This includes providing advices and opinion on project technical challenges, solutions and choices, with a focus, at this stage, on the project's architecture.

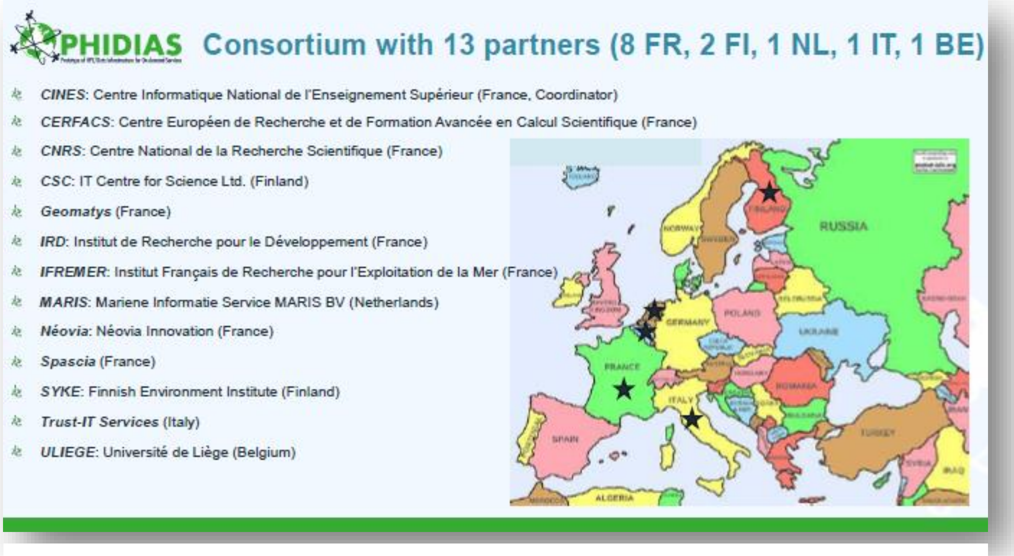
3.2.3 Overview of PHIDIAS

In 3 years (from 01/09/2019 to 01/09/2022) the PHIDIAS project aims to build a prototype for Data/High Performance Computing (HPC) services based on Earth sciences cases. In this respect, the consortium will develop and provide new services to discover, manage and process spatial and environmental data produced by research communities tackling scientific challenges such as atmospheric, marine and earth observation issues. One of the particularities of this project is that it is also focused on services' development (as required by the call), whereas the regular CEF projects are more infrastructure's focused.

Co-financed by the Connecting Europe Facility of the European Union

The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

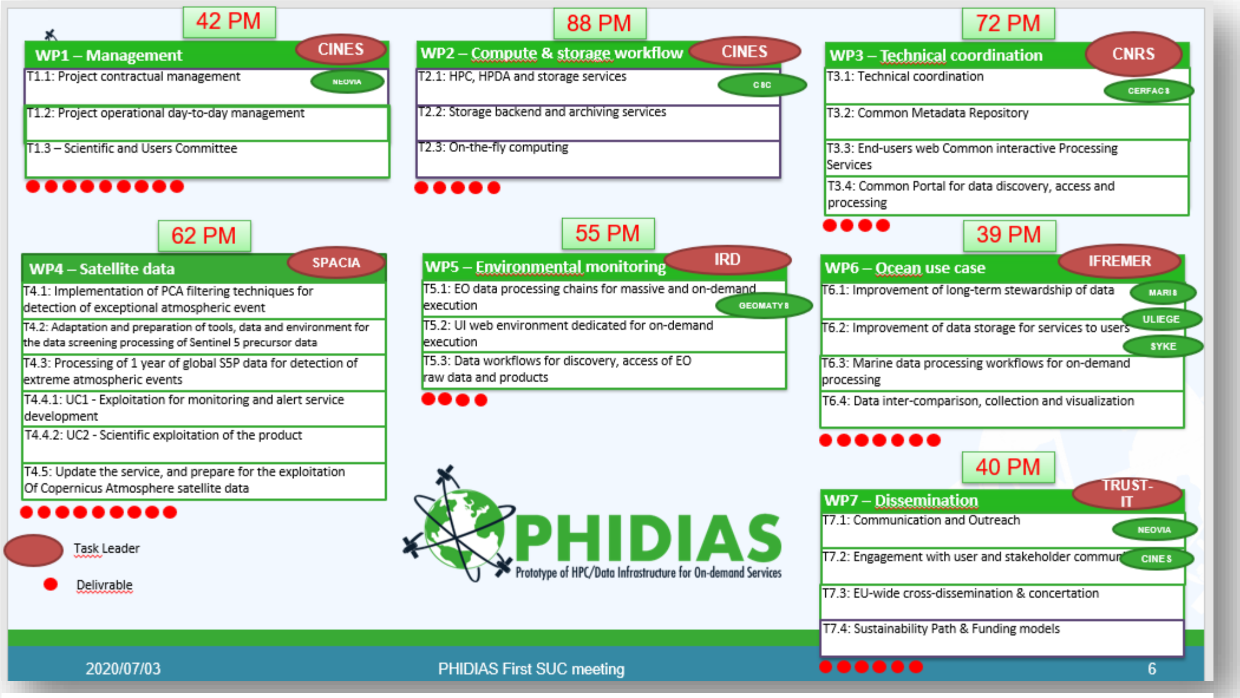
The project has a global budget of 3 519 476€, with a CEF Grant of 75% of the total budget.



PHIDIAS Consortium with 13 partners (8 FR, 2 FI, 1 NL, 1 IT, 1 BE)

- *CINES*: Centre Informatique National de l'Enseignement Supérieur (France, Coordinator)
- *CERFACS*: Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique (France)
- *CNRS*: Centre National de la Recherche Scientifique (France)
- *CSC*: IT Centre for Science Ltd. (Finland)
- *Geomatys* (France)
- *IRD*: Institut de Recherche pour le Développement (France)
- *IFREMER*: Institut Français de Recherche pour l'Exploitation de la Mer (France)
- *MARIS*: Mariene Informatie Service MARIS BV (Netherlands)
- *Néovia*: Néovia Innovation (France)
- *Spascia* (France)
- *SYKE*: Finnish Environment Institute (Finland)
- *Trust-IT Services* (Italy)
- *ULIEGE*: Université de Liège (Belgium)

Figure 2 - The PHIDIAS Consortium



42 PM
WP1 – Management — CINES
T1.1: Project contractual management — NEOVIA
T1.2: Project operational day-to-day management
T1.3 – Scientific and Users Committee
●●●●●●●●●

88 PM
WP2 – Compute & storage workflow — CINES
T2.1: HPC, HPDA and storage services — CSC
T2.2: Storage backend and archiving services
T2.3: On-the-fly computing
●●●●●

72 PM
WP3 – Technical coordination — CNRS
T3.1: Technical coordination — CERFACS
T3.2: Common Metadata Repository
T3.3: End-users web Common interactive Processing Services
T3.4: Common Portal for data discovery, access and processing
●●●●

62 PM
WP4 – Satellite data — SPACIA
T4.1: Implementation of PCA filtering techniques for detection of exceptional atmospheric event
T4.2: Adaptation and preparation of tools, data and environment for the data screening processing of Sentinel 5 precursor data
T4.3: Processing of 1 year of global S5P data for detection of extreme atmospheric events
T4.4.1: UC1 - Exploitation for monitoring and alert service development
T4.4.2: UC2 - Scientific exploitation of the product
T4.5: Update the service, and prepare for the exploitation Of Copernicus Atmosphere satellite data
●●●●●●●●●

55 PM
WP5 – Environmental monitoring — IRD
T5.1: EO data processing chains for massive and on-demand execution — GEOMATYS
T5.2: UI web environment dedicated for on-demand execution
T5.3: Data workflows for discovery, access of EO raw data and products
●●●●

39 PM
WP6 – Ocean use case — IFREMER
T6.1: Improvement of long-term stewardship of data — MARIS
T6.2: Improvement of data storage for services to users — ULIEGE — SYKE
T6.3: Marine data processing workflows for on-demand processing
T6.4: Data inter-comparison, collection and visualization
●●●●●●●

40 PM
WP7 – Dissemination — TRUST-IT
T7.1: Communication and Outreach — NEOVIA
T7.2: Engagement with user and stakeholder commun — CINES
T7.3: EU-wide cross-dissemination & concertation
T7.4: Sustainability Path & Funding models
●●●●●●

Task Leader
● Delivrable

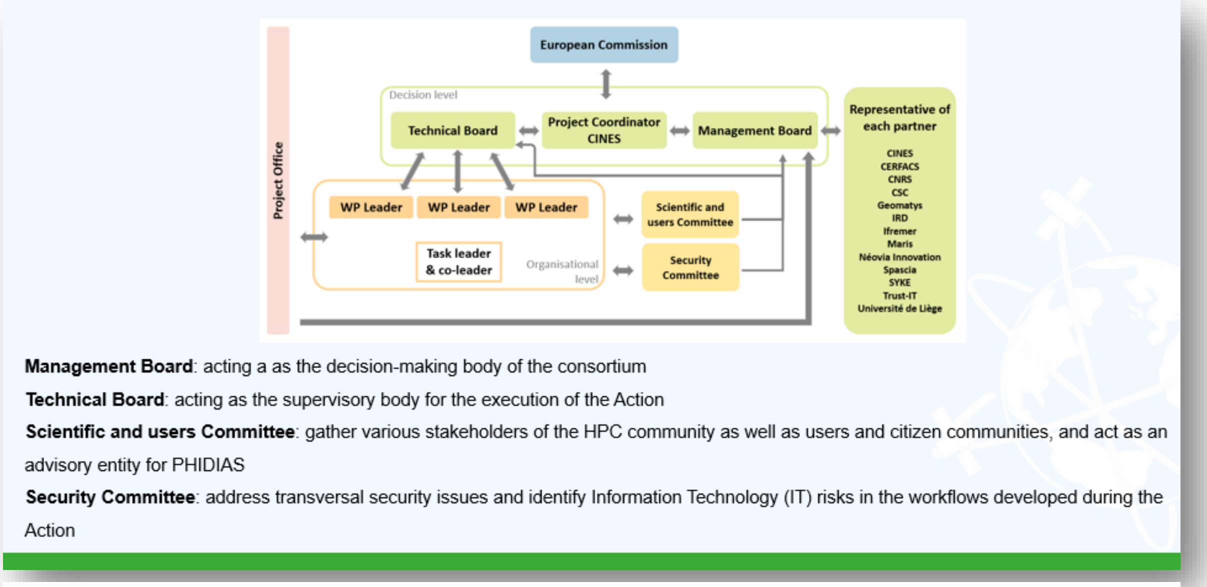
PHIDIAS
Prototype of HPC/Data Infrastructure for On-demand Services

2020/07/03 PHIDIAS First SUC meeting 6

Figure 3 - PHIDIAS WPs & tasks

Co-financed by the Connecting Europe Facility of the European Union

The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.



Management Board: acting a as the decision-making body of the consortium

Technical Board: acting as the supervisory body for the execution of the Action

Scientific and users Committee: gather various stakeholders of the HPC community as well as users and citizen communities, and act as an advisory entity for PHIDIAS

Security Committee: address transversal security issues and identify Information Technology (IT) risks in the workflows developed during the Action

Figure 4 - PHIDIAS organisational and decision's structure

The consortium relies on one central collaborative tool for the life of the project: Confluence.

There are 4 main challenges tackled by PHIDIAS:

- **[Challenge #1]** Handling the diversity of data coming from the Earth System Research Infrastructure.
- **[Challenge #2]** Developing and testing transversal methods and tools that can be applied to data coming from other scientific domains such as health and environment data.
- [**Challenge #3**] Scalability of Data processing tools by allowing the proper transfer of processing running on small machines and clusters to supercomputers.
- [**Challenge #4**] Industrialization and strength development of HPC/HPDA/AI workflows to touch and commit communities beyond PHIDIAS.

The main data provider of the PHIDIAS project is Data Terra research infrastructure, comprised of 4 pillars (AERIS for the atmospheric data, ODATIS for oceanic data, FORM@TER for earth data and THEIA for continental surfaces's data). Data Terra will provide the 3 scientific use cases composing the PHIDIAS project:

- **[Use case #1]** Intelligent screening of large amount of satellite data for detection and identification of anomalous atmospheric composition events (WP4, leader: SPASCIA) **→ AERIS**

Co-financed by the Connecting Europe Facility of the European Union

The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

- **[Use case #2]** Big data Earth Observations: processing on-demand for environmental monitoring (WP5, leader: IRD) → **THEIA**
- **[Use case #3]** Ocean (WP6, leader: Ifremer) → **ODATIS**

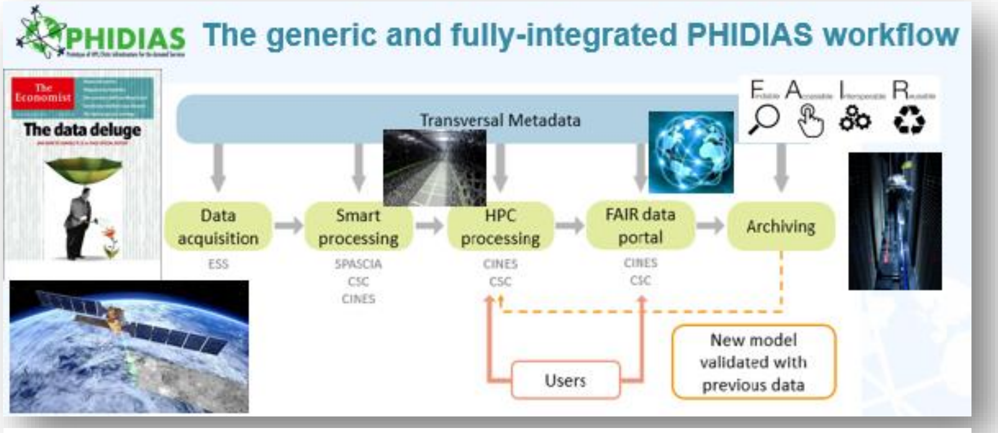


Figure 5 - Synthetic view and components of the main PHIDIAS workflow

All these use cases need an access to a supercomputer's infrastructure, in conformity with the FAIR paradigm. Furthermore, an archiving of the data will be created on the CINES platform.

PHIDIAS project shares similarity to the former Herbadrop/Icedig project[1], based on the herbaria use case, which made use of the OCR recognition pattern and an elastic search engine. The principal difference between the two projects is that PHIDIAS uses satellite data and may enrich the catalogue of services proposed to the scientific communities.

Questions & Answers:

Q1 (Jean-Pierre Vilotte): The challenge #4 of the project (industrialisation and engaging other communities) is wide. What kind of other communities will the project touch and how will its solutions and services adapt to these communities – for example with biology or HS?

A1: The WP7 has a subtask dedicated to "sustainability path & funding model", notably because the MESRI had asked to work on the cost modelling at the end (process' costs on HPC/Data centre, costs of "generic" workflows, etc.).

Q1bis (Jean-Pierre Vilotte): Funding models are part of the answer but also what is developed by PHIDIAS to fit with specific users' technical needs – for example: will the PHIDIAS' architecture be reusable by other scientific communities?

[1] Results can be consulted at : https://opendata.cines.fr/

A1bis (Richard Moreno): The workflow is very generic. It moves data to HPC centres using iRODS. Using standards is also a key component in making the project's results reusable. Regarding the reaching of other communities, PHIDIAS is, for example, already working with Humanum through iRODS for transfers of data.

Q2 (Jean-Pierre Vilotte): On the other part of your workflow, you will have to deal with different kind of data: How will you deal with it and what level of data will you be diffusing (i.e. data or advanced science product, both, etc.)?

A2 (Richard Moreno): The workflow is focused on geographical data (*in situ* and satellite), using a data model tailored for environmental/geographical data. It is therefore not possible to use it directly for another community's type of data, but the workflow can be adapted (as there are some standards related to earth sciences).

A2bis (Pascal Prunet): The project is working with pilot users. The second part of the project will be dedicated to verifying that the services are useful to scientific users, including companies (ex for WP4). About the level: both are used. For WP4 for example, there are level 1 data (on the atmospheric spectrum, directly for scientific users) and level 2 (atmospheric gazes scrutinized to detect extreme events).

Q3 (Sylvie Joussaume): How does it differ for usual CNES process for satellite data? How much this is different of the HPC culture of CNES?

A3 (Richard Moreno): Though interested and following the outcomes of the project, the CNES is not a partner of the project. The CNES will especially follow results on solutions explored to deal with a large variety of data. At the moment, handling of space data at the CNES is done on a project per project basis but there is nothing scheduled to cross and compare data on more general missions which would be possible with PHIDIAS' results.
PHIDIAS will allow the users to use data from different missions and themes with a common meta data model, a common services' catalogue, etc.
CNES intends to reuse the outputs of PHIDIAS to develop such a service on a central data lake: The data model by PHIDIAS is the basis for that goal. Finally, CNES has an interest in PHIDIAS through their links with Data Terra.

3.2.4 Ocean use case

As an evolving ecosystem, ocean is a difficult ecosystem to observe (especially deep-sea areas). Its observation requires several complementary systems (from satellite to underwater vehicles, etc.) and depend on inter-comparisons and co-processing of all produced data.

The goals of this use case are:

- Combine and collocate data from several data sources (in situ & satellite)
    - Enhancing data archiving (most observations cannot be reproduced) to facilitate data reuse

    - Facilitate and speed up co-localisation and process of data from different sources
- Adopting new data structures (based on big-data technologies)
    - DataCubes
    - NoSQL databases (numerical data): Cassandra, MongoDB, etc.
    - Semantic Web (text data)
- Providing on demand data browsing and processing facilities

The use case focuses its work on two geographical areas (case-studies):

- Surface Salinity in North Atlantic
    - CTD (SeaDataNet),
    - Argo Floats (CMEMS),
    - SMOS satellite.
- Chlorophyll in North-East Atlantic and Baltic Sea
    - CTD and bottles (SeaDataNet),
    - BGC Argo floats (ARGO GDAC),
    - Ferrybox,
    - Sentinel 2 images (DIAS WEkEO)

Improvement of Data Storage:
The present data structures (e.g. collections of netCDF files) of *in-situ* marine data are not very efficient, due to the large number of files and/or the heterogeneity of data observations. The challenges addressed are the following ones:

- Access quickly to a few number of observations within a large number of observations: data visualization and data selections on web portals
- Access to large number of observations within a large number of observations: data processing or parallel processing (e.g. machine learning algorithms, interpolation software such as DIVAnd, etc.)

The solutions in development are:

- Data visualization: test of No-SQL databases (e.g. Cassandra - cassandra.apache.org)
- Data processing: test of "data cubes" structures such as Parquet (columnar storage format from the Haddoop ecosystem - parquet.apache.org)

About the "on demand Data Processing", the PHIDIAS project develops a "Virtual Research Environment" dedicated for users, allowing:

- Jupyter Notebook as a basis to develop workflows
    - Annotate, compare, conserve and share expertise
    - Access to different data structures in one environment (access to specified data on premise or remotely)
- Scripting in various languages (Python, R, Julia, etc.)
- Access to Pangeo components
- Use of DIVAnd software
    - Software for gridding data, using a finite-element method

  - Developed in Julia programming language
  - Extension for satellite images
- GIS features
  - Visualisation of inputs (study area), outputs (compare results of different processes) and images
  - Existing tools: DIVAnd online, Sextant, Geomatys, etc.

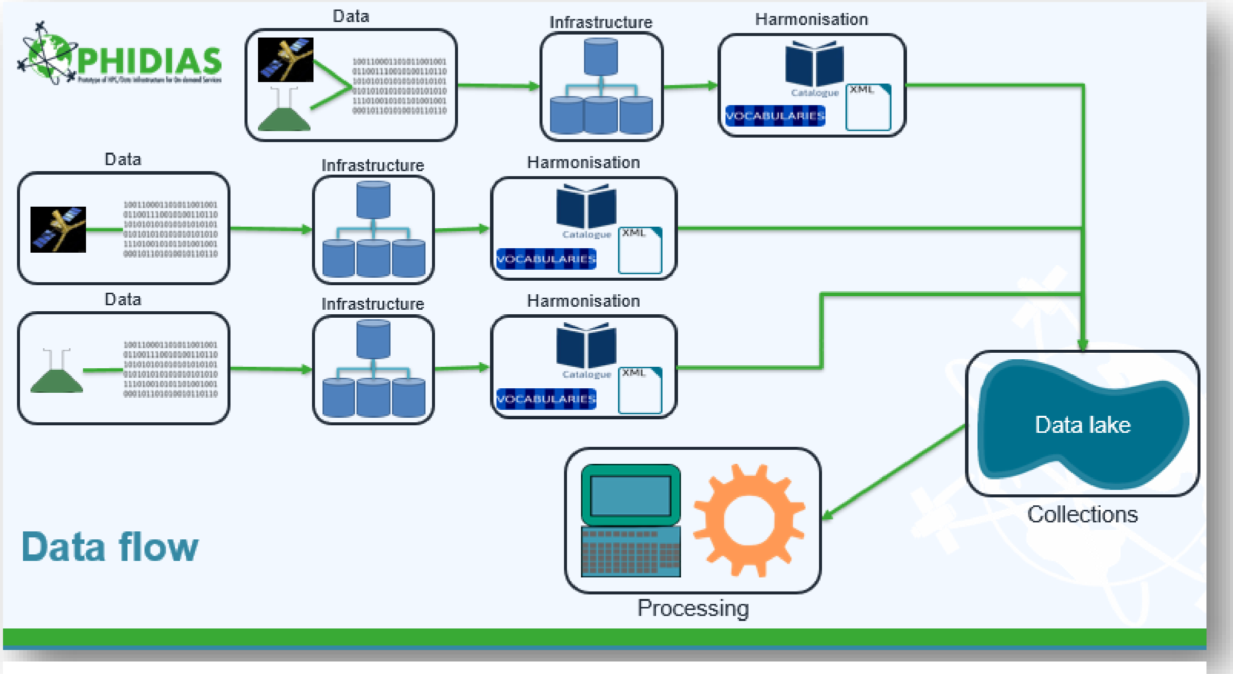


Figure 6 - Data flow description of the PHIDIAS project

__Questions & Answers:__

Q1 (Bruno Raffin): Is the data process continuous, e.g. in streaming mode or triggering a “one-shot” workflow?

A1 (Gilbert Maudire): It is focused for on-demand processing but not with a continued workflow. The goal is to provide to users several platforms to process data on specific area and time slots. It is just a part of the stream to merge in the data lake from different sources but not on the upstream processing.

A1bis (Boris Dintrans): HPC centres in France do not have the resources for “on the fly” computing as there are not in production mode like Météo France or CNES. HPC centres are currently organized with a batch system and resources are shared between users. There is no machine to machine workflow in production mode for HPC centres now.

Co-financed by the Connecting Europe Facility of the European Union

The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

Q2 (Bruno Raffin): Though the project makes use of HPC resources; the Ocean use case seems more big data oriented. What is the role of HPC for this use case?

A2 (Gilbert Maudire): Indeed, the use case is more linked with HPDA and is less HPC-oriented but still relies on the use of HPC to answer the difficulties for processing big amount of data for the screening of large areas (North Atlantic Ocean for example). For this purpose, HPC resources are critical to improve and speed up this processing using DIVAnd.

Q3 (Jean Pierre Vilotte): As the services will need to access HPC resources at certain stage of the data processing, huge amount of data will be required to be moved. How will the project deal with such critical data logistics and avoid bottlenecks? Especially how the project will use containerisation and virtualisation's possibilities?

A3 (Gilbert Maudire): The general philosophy is to minimize the movement of data. If there is move of data from site to site (when HPC's computing is inevitable), the goal is to optimize it with data technologies like iRODS. Furthermore, PHIDIAS takes benefit of containerisation (for example DIVAnd too can be deployed as Docker or Singularity container) and makes it work in HPC environment.
HPC is also one motivation to move or duplicate the data. In some cases, data sets are not so large but very numerous. Coming from one data site with a lot of files to another with a filter to reduce the number of files in treatment is one option. Some duplication will be accepted especially for *in situ* processing.

A3bis (Jean-Christophe Penalva): The general philosophy is first to set-up the architecture to minimize the needs for data movement – which implies to keep big data on site and apply as much computing as possible close to the data location.
When data movement is inevitable, the use of iRods will make it transparent from site to site. Containerisation is possible and already implements both at CINES and CSC.

Some precisions on this last question will be given in the part "Infrastructure and overall architecture elements".

3.2.5 Big data earth observations use case

The objective of WP5 is to merge scientific experimental algorithms and catalogues with the communities' user-needs.

The technical sub-objectives are the following ones:

- Dedicated environment adapted to the target users
- Ability to produce maps over large areas in a systematic manner
- Open the dissemination of processing chains outputs in FAIR way
- Improvement of data reusability in perspective of EOSC
- Leveraging AI techniques to provide alternative image classification methodologies

Co-financed by the Connecting Europe Facility of the European Union
The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

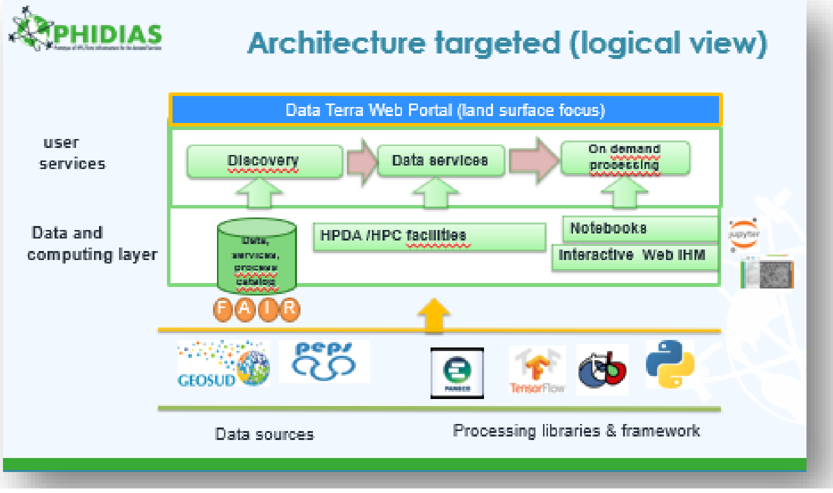
There are two Case studies we want to develop in the WP5.

The first one is focused on Sentinel-1/Sentinel-2-derived Soil Moisture product at Plot scale (S2MP) over agricultural areas.

- The point is to ensures better monitoring of the soil moisture in agricultural areas for the two target users: Scientific and farming sector. The monitoring relies on the free access to Sentinel-1 (SAR satellite) and Sentinel-2 (optical high resolution 10x10 m) with high revisit to model water cycle (for the scientific users) and map the irrigation activities (for the farming sector)
- These issues tackled by PHIDIAS are the interactive on-demand HPC processing on specific zone and a large temporal depth and the Big data access and data outputs FAIRness.

The second case study is about the remote sensing images processing with artificial intelligence: application to land cover mapping and super-resolution using both Sentinel-2 high-resolution images and semantic segmentation. (SPOT & Sentinel-2).

- PHIDIAS addresses the double issue of the HPC's GPU architecture to test scalability of IA approach over national territory and Big data access and data outputs FAIRness.



PHIDIAS

Architecture targeted (logical view)

Data Terra Web Portal (land surface focus)

user services

Discovery

Data services

On demand processing

Data and computing layer

Data, services, process catalog

HPDA /HPC facilities

Notebooks

Interactive Web IHM

Data sources

Processing libraries & framework

Figure 7- Architecture targeted data flow description of the PHIDIAS project

Questions & Answers:

Q1 (Sylvie Joussaume): What type of HPC resources (Tier-1 or Tier-2) do the three use cases?

A1 (Jean-Christophe Desconnets): There are two modes of production. The need of HPC infrastructure is systematic for very large territory, as the communities do not have access to such resources. For on-demand problem, it would more be a matter of allocation modes to cope with resources needed for each iteration. Offers coming from HPC and data centres do

Co-financed by the Connecting Europe Facility of the European Union

The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

not provide easy, agile, flexible access to the infrastructure and this is a major bottleneck to industrialise HPC towards the community that PHIDIAS aims to solve.

Q2 (Jean-Pierre Vilotte): The on-demand computing could mean a lot of different things including urgent computing: can it be clarified? For instance, it can either stress the capacity or availability of the HPC infrastructure which would require two different types of management for the HPC infrastructure. How do you provide that easy access to the HPC centres?

A2 (Jean-Christophe Desconnets): PHIDIAS does not address urgent computing but only on-demand computing. The need that is covered is to test, configure and/or execute computation for a specific study in the quickest time.

Comment on A2 (François Bodin): The addressed problem sounds more like a capacity problem.

Q3 (Bruno Raffin): Is the processing part of the project?

A3 Not really. There is already some processing at the right level of developing. The principal component of processing is to select the data (outliers are detecting relevant indicators). The risk lies on the interpretation stemming from processing, not the processing itself.

3.2.6 Satellite data use case

From 2021, European atmospheric sounding missions will deliver each day several terabytes of raw datacubes at high spatial/temporal/spectral resolutions. This represents an unprecedented amount of atmospheric data, with improved quality and coverage.
Therefore, the key objective is to provide the capacity of intelligent screening of large amounts of satellite data for targeting scenes or events of interest in view of their dedicated processing or exploitation; by the cross use of HPC and HPDA (high-performance data analytics).

This WP proposes to develop, test and produce two prototypes of the approach with Sentinel 5 Precursor (S5P) data/products (Sentinel 5 Precursor is in operation since April 2018):

- 1. PCA based screening of L1 data (SWIR) for detection of extreme events. Based on experiences and methods developed for IASI, implementation and consolidation of algorithms and tools for generic processing of atmospheric spectra recorded by S5P.
- 2. New AI methods for objective/automatic detection of plumes from L2 products (CO, and eventually CH_4 NO_2, SO_2): SPASCIA experience with physical methods analysing specific signal enhancements

Co-financed by the Connecting Europe Facility of the European Union
The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

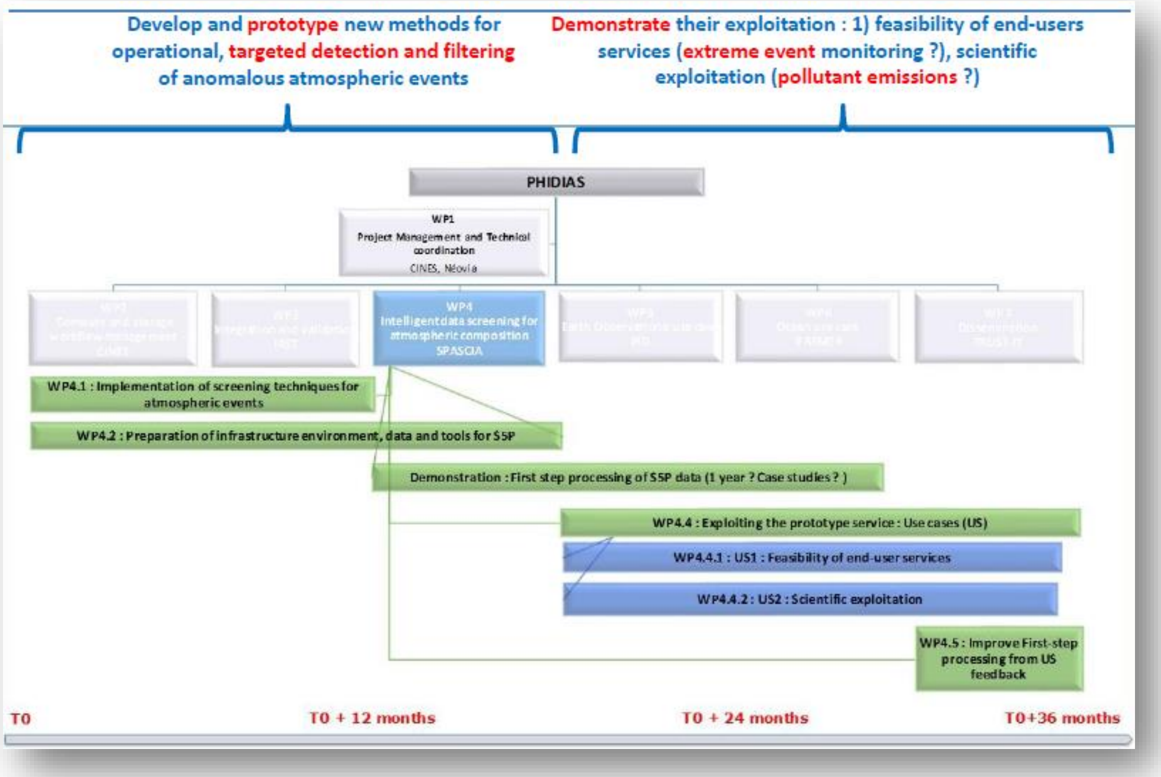

Develop and prototype new methods for operational, targeted detection and filtering of anomalous atmospheric events
Demonstrate their exploitation : 1) feasibility of end-users services (extreme event monitoring ?), scientific exploitation (pollutant emissions ?)
PHIDIAS
WP1
Project Management and Technical coordination
CINES, Néovia
WP4
Intelligent data screening for atmospheric composition
SPASCIA
WP4.1 : Implementation of screening techniques for atmospheric events
WP4.2 : Preparation of infrastructure environment, data and tools for S5P
Demonstration : First step processing of S5P data (1 year ? Case studies ?)
WP4.4 : Exploiting the prototype service : Use cases (US)
WP4.4.1 : US1 : Feasibility of end-user services
WP4.4.2 : US2 : Scientific exploitation
WP4.5 : Improve First-step processing from US feedback
T0
T0 + 12 months
T0 + 24 months
T0+36 months

Figure 8 – Development plan of WP4 use cases and services

The Covid-19 crisis was the occasion to assess the impact of the human activity reduction on air pollution by comparing the first 4 months periods of 2019 and 2020 on a daily, weekly and monthly basis for 4 major cities in Europe. It appeared that reductions in the pollution level (using NO_2 tropospheric column as a proxy) have been observed from Mid-March and for April 2020 (52% +/- 9% for Paris; 28% +/- 8% for Milan region; 54% +/- 16% for Madrid; not significantly observed for Athens), as compared to the same periods in 2019.

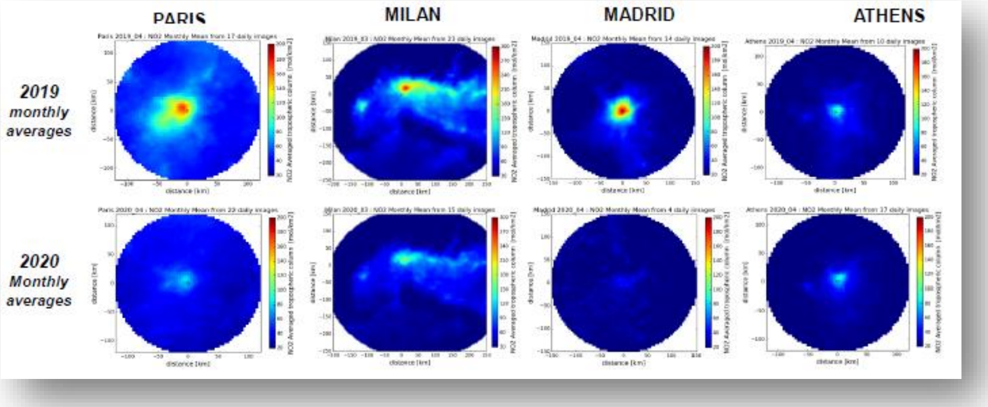


Figure 9 - PHIDIAS assessment of NO2 concentration during Covid-19 confinement in Europe in April 2020

The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

Question & Answers:

Q1 (Bruno Raffin): Regarding the air pollutant assessment, are the Particulate Matter (PM) directly identified or derivated from the available data?

A1 (Pascal Prunet): The monitoring done was focused on gaz, but PM products exist with ICARE based on S5P data.

Q2 (Jean-Pierre Vilotte): What would be the difficulties to integrate uncertainties such as bias in using IA?

A3 (Pascal Prunet): There is no full answer at this stage. There are two different processings: First is to select some data, integrating data noises and building outliers using specific indicators as thresholds. The data are selected to exclude the noise.
For IA, it cannot be answered yet. This is not a classical IA processing though based on learning, but final processing will be new: It takes into account noise, but the key question is how to transform the noise and deliver a product that includes uncertainties… This issue shall be discussed more in depth.

Q4 (Bruno Raffin): Is developing AI also a part of the project?

A4 (Pascal Prunet): No, the development of AI for detection of extreme events is done in parallel and based on already demonstrating capacity developed for the biology sector that is reworked for space data. There is a risk here as we do not know if the transformation will be efficient.

Q5 (Bruno Raffin): Will this use case work on stream processing?

A5 (Pascal Prunet): Yes, we will try to deal with that point in PHIDIAS.

Q6 (Bruno Raffin): How do you factorize the development of the base layers of the workflow, which is different from other WPs, and how much it is shared between the WPs?

A6: Such matters are covered by WP3 and shall be answered by Richard Moreno at the end of the last presentation.

Co-financed by the Connecting Europe Facility of the European Union

The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

3.2.7 *Infrastructure and overall architecture elements*

The definition of the PHIDIAS architecture is a crucial collaborative work of the WP2 involving all partner form WP3 to WP6.

The current works revolve around the following points:

- Data distributed on different sites: Each site continues to produce data and the data is shared.
- The WP3 is working to make the heterogeneous Formats more interoperable
- Unilateral treatment:
  - Referred treatment
  - Database of available treatments
  - Multi-site workflow capability

Some difficulties arose during the work on the infrastructure. A difficulty of access to the data catalogue (business catalogue) and a difficulty regarding satellite's data transfers have been identified. A work with IRODS is made to allow transparent view of the data with micro-services to make asynchronous treatment of the data and to manage and find the better way to optimize the data transfers' moment. Furthermore, it must be recalled that the access rights are managed by each producer/owner, nonetheless, there are some issues regarding these results access (public or restricted).

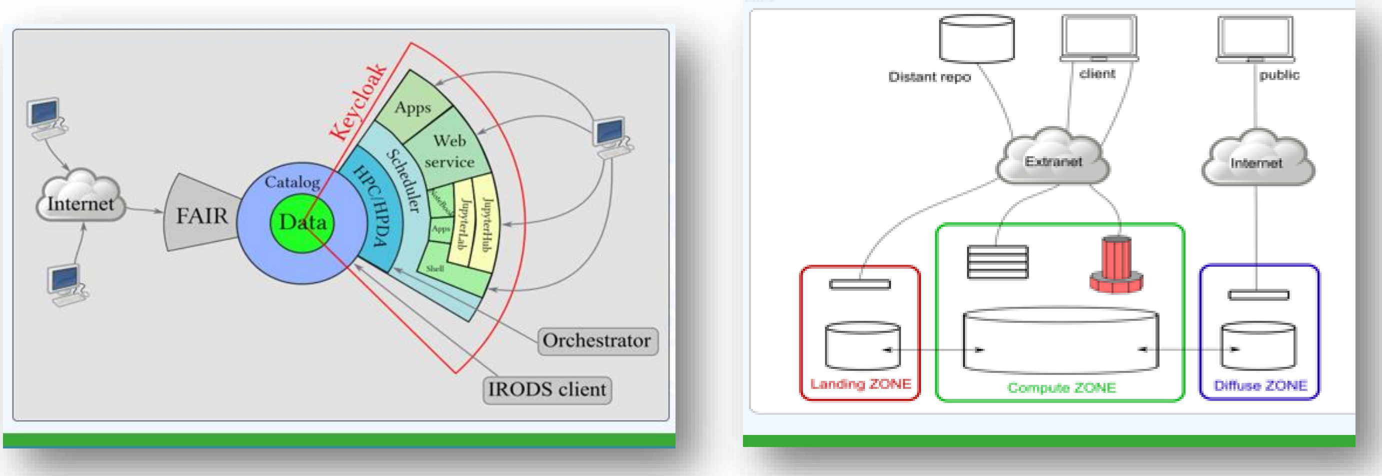


Figure 10 - Current version of PHIDIAS Architecture (details)

The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

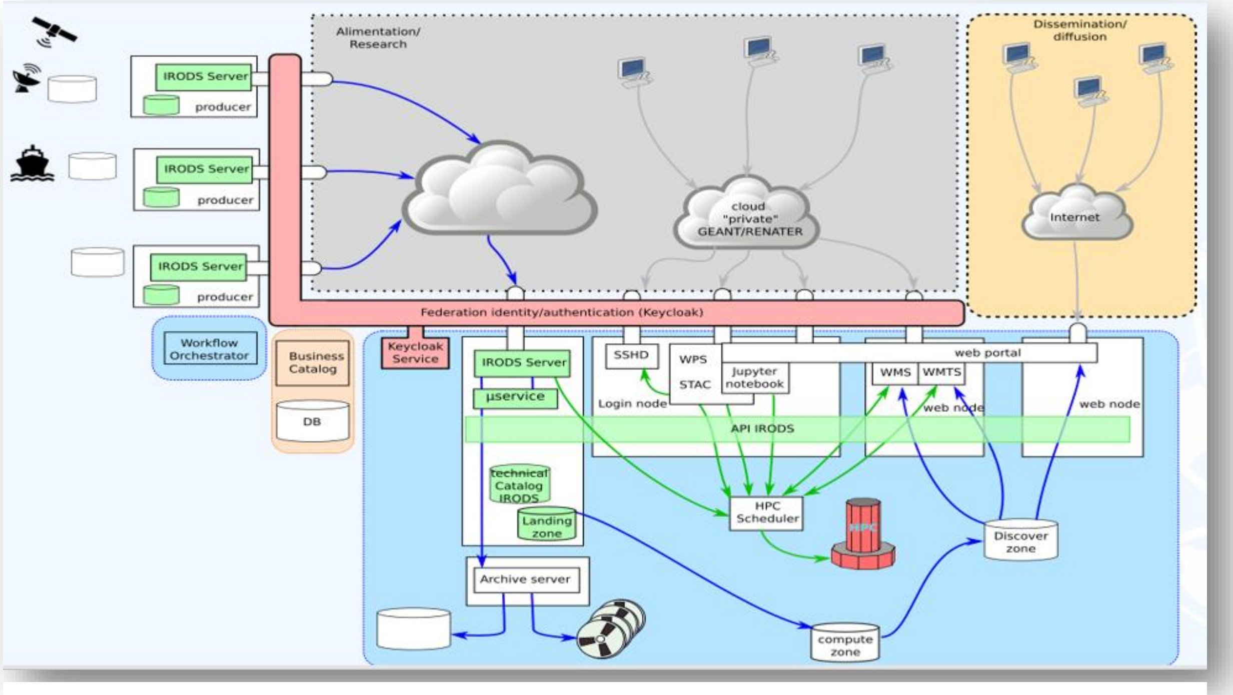


Figure 11 - Current version of PHIDIAS architecture (global)

Question & Answers:

Q1 (Michele Fichaux): About the interoperable formats, which principle to follow to achieve their definition? Will you define generic formats or rules to be followed by data producers?

Q2 (Jean-Pierre Vilotte): PHIDIAS is very ambitious as it touches many things (infrastructure, data, etc.) and then needs a lot of tools, to define identification resolution protocol, etc.
How much will the project develop or use what is done in many other projects, especially EOSC. Thus, how is PHIDIAS interoperable among different other initiatives – which can improve its sustainability?

A1 & 2 (Richard Moreno): What PHIDIAS developed is based on standards used by the concerned communities. The core of the metadata model is based on ISO (e.g. native aligned with INSPIRE) and our users are using this model. The project will add other interoperable result (for instance OpenSearch, STAC which is a new standard completely compatible with Cloud computing and Pangeo, etc.). Furthermore, it is linked with Data Terra (+ integrate cloud computing, with France grid for instance)
The link with EOSC is made by three channels key partners are already active in EOSC-Pillar, plus WP3 team is involved in EnvriFair and lead one implementation network of GoFair.

Co-financed by the Connecting Europe Facility of the European Union
The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

3.2.8 *Common system*

The objectives of the WP3 is to assure the technical coordination of WP4, WP5 and WP6, build a system to remove existing data silos and allow user to develop programs allowing to handle the full "Earth System" perimeter (e.g. programs in the field of Climate or Sustainable development).

It has 4 tasks:

- **Task 3.1**: Technical coordination (M1-M36; IRST)
- **Task 3.2**: Common Metadata Repository (M1-M18; IRST, Geomatys, IRD, CERFACS)
- **Task 3.3**: End-users web Common interactive Processing Service (M18-M32; IRST, with implication of Geomatys, IRD, IFREMER, CERFACS)
- **Task 3.4**: Common Portal for data discovery, access and processing (M18-M32; IRST, with implication of Geomatys, IRD, IFREMER, CERFACS)

The task 3.1 (Technical coordination) is about the work on architecture with WP4, 5 & 6; the WP2 managing the global architecture.

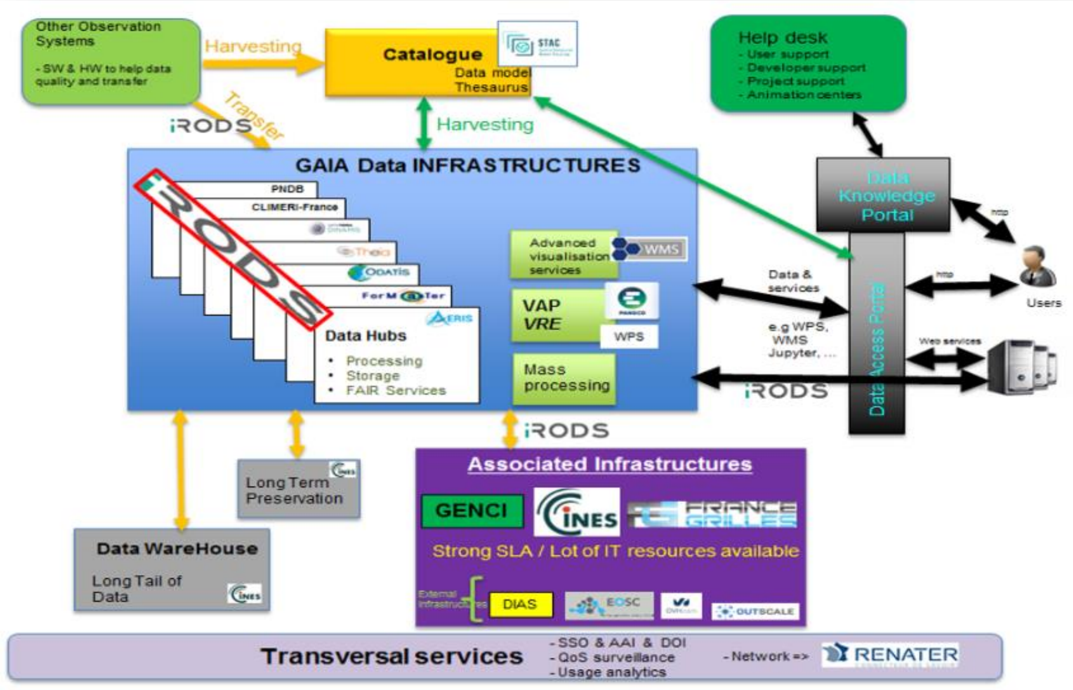


Figure 12 - General architecture of GAIA Data

The task 3.2, is dedicated to developing ways to remove existing data silos (e.g. between the different themes as Ocean, Atmosphere or Land Surface):

- To choose the most appropriate tool to build this Common Metadata Repository, based on experience of the consortium (e.g. Geomatys) and of the Data Terra RI

The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

- To define the Common Information Model and the common vocabulary allowing to discover, access and process the data from WP4, WP5 and WP6; based on work done in the context of Data Terra. Furthermore, the goal is to develop, install, and exploit this Common Metadata Model on the CINES infrastructure.

The current works revolves around the making of a first version of the Common information Model that will be issued in the 3rd Trimester and a first mock-up of Common Metadata Repository (issue for the same date).

The Task 3.3 (End-users web Common interactive Processing Service) is based on the Common Interface Document ([D.3.1.1]) and existing open source frameworks (e.g. Pangeo or Open DataCube). It will allow the creation of an interactive prototyping graphic environment (e.g. notebook). Finally, some work will be done on the current data and processing access portal enhancement to allow the discovery of raw data catalogues (from WP4, WP5 and WP6) and products derived from processing.

Some solutions have been analyzed and should be further explored:

- WPS on-demand services (which will be produced by WP5)
- VRE (Virtual Research Environment): VIP, an initiative from France-Grilles that will be analyzed in the context of WP5 and GALAXY, which is mainly used by the Biodiversity community.
- VAP (Virtual Analysis Platform): Some analysis and work around AI4GEO VAP (based on PANGEO) have been conducted.

The Task 3.4 (Common Portal for data discovery, access and processing) is dedicated to the development of a WWW portal allowing to discover, access and process the data from WP4, WP5 and WP6. It will be built upon the Common Metadata Repository and the End-users web Common interactive Processing Service.
This portal will be nurtured of the experiences from existing portals developed within current and past European projects and initiatives such as EOSC, Copernicus and IS-ENES). Furthermore, it will be developed in compliance with FAIR and INSPIRE principles and shall be interoperable with other portals and generic services (EUDAT, EOSC, IS-ENES).
A first mock-up is foreseen for the 3rd Trimester.

Additional activities are comprised of tests between CNES & CINES (for WP5) on grid of data and services using iRODS and the installation in CINES of a repository for “Long Tail data” with Dataverse.

Co-financed by the Connecting Europe Facility of the European Union

The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

3.2.9 Discussion

As required by some external experts to allow better comprehension and guidance, the PHIDIAS Confluence Space will be open to the SUC's members: access logs will be provided by PHIDIAS' PMO.

The minutes will be prepared and shared: the potential following questions from it will be answered by email.

3.2.10 SUC President election process and next meeting

The election of the SUC President was supposed to be done during the meeting. Because of the deep and long discussions which arose during the meeting, it will be organised later by email. By default, if there are no applications, François Bodin will accept the role.

Edit: As no other member of the SUC declared to be candidate for the president's position, François Bodin became president of the Scientific and Users Committee on the 28th of July 2020.

The next meeting will be held in December. It will focus on 3 major points: technical subjects (in-depth presentation of the PHIDIAS architecture included), ecosystem subject and also the presentations, from external experts, of relevant external projects (as the PHIDIAS project could be connected to other European projects and needs a "macro-vision").

4 Annexes

4.1 *Annex 1 - List of the 22 participants and screenshot of the meeting*

4.2 *Annex 2 - List of presentations (available at Confluence)*

- PHIDIAS overview – an authorised access is needed
- WP2 presentation – an authorised access is needed
- WP3 presentation – an authorised access is needed
- WP4 presentation – an authorised access is needed
- WP5 presentation – an authorised access is needed
- WP6 presentation – an authorised access is needed

4.3 *PHIDIAS Scientific and Users Committee – Coordinator selection process*

The Phidias project includes a Scientific and Users Committee, which gathers various stakeholders of the HPC community as well as users and citizen communities. The role of this committee is to act as an advisory entity for PHIDIAS. A specific note describe the role and operation of this committee.

This committee is coordinated by a Committee Coordinator whose role is to ensure the coordination and logistic of the committee and interact with the project in particular with the Project Coordinator, the Project Management Office, the Management Board and the

Co-financed by the Connecting Europe Facility of the European Union

The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

Technical Board. The Committee Coordinator is member of the Scientific and User Committee.

At the beginning of the project, the Project Coordinator CINES and Neovia Innovation looked for persons outside the project able to launch and coordinate the Scientific and User Committee and a list of 15 persons have been established.

Then the following five criteria have been considered:

- Recognized scientific and technical experience in the field of Data acquisition and transfer, Scientific Computing, HPC and IoT
- Understanding, knowledge and connections to the world-wide context and community in Europe, Asia and USA
- Previous experience of Research and Innovative European projects
- Connection and knowledge of the European Commission in the Digital field and to associated organisations such as EuroHPC, EXDCI, ETP4HPC, BDVA, HiPEAC…
- Availability

Each candidate has then been evaluated on each criterion by CINES (Boris Dintrans) and Neovia Innovation (Thierry Bidot) and a ranking done.

The notes were for each criterion 1=low; 2=medium; 3=high

The person best ranked is François Bodin from U. Rennes 1 who contacted by the project accepted the role and signed a subcontract with Neovia Innovation.

Nom	Organisation	Recognized scientific and technical experience in the field of Data acquisition and transfer, Scientific Computing, HPC and IoT	Understanding, knowledge and connections to the world-wide context and community in Europe, Asia and USA	Previous experience of Research and Innovative European projects	Connection and knowledge of the European Commission in the Digital field and to associated organisations such as EuroHPC, EXDCI, ETP4HPC, BDVA, HiPEAC…	Availibilty		Overall evaluation
Alain Refloch	ONERA	Medium	Low	High	Low	Low		8
Antonio Parodi	CIMA	Medium	Low	High	Low	Medium		9
Catherine Lambert	CERFACS	Low	Medium	High	High	Low		10
Thomas Lippert	JUELICH	Medium	Medium	Medium	High	Low		10
Sergi Girona	BSC	Low	High	High	High	Low		11
Thomas Shultess	ETH ZURICH	Medium	High	Medium	High	Low		11
Uli Ruede	U. ERLANGEN	Medium	Medium	High	Medium	Medium		11
Wolfang Nagel	U. DRESDEN	Medium	Medium	High	High	Low		11
Alison Kennedy	STFC	Medium	Medium	High	High	Low		11
Bernd Mohr	JUELICH	Medium	High	High	High	Low		12
Jean-Claude André		Medium	Medium	High	High	Medium		12
Thomas Fahringer	U. INNSBRUK	Medium	Medium	High	High	Medium		12
François Bodin	U RENNES 1	High	High	High	High	Medium		14

Figure 13 - List of candidates for the position of PHIDIAS SUC Coordinator

Co-financed by the Connecting Europe Facility of the European Union

The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.